REPORT DOCUMENTATION PAGE

Approved

AFRL-SR-AR-TR-04-

0023

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED
11112003	Final	1/01/01 - 12/01/02

4. TITLE AND SUBTITLE
Steganalysis of Digital Watermarking Techniques

5a. CONTRACT NUMBER

5b. GRANT NUMBER
F49620-01-1-0243

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Nasir Memon
Jessica Fridrich

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Polytechnic University
Brooklyn, NY 11201

SUNY Binghampton
Binghampton, NY 13902

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Department of the Air Force
Air Force Office of Scientific Research
4015 Wilson Blvd.
Arlington, VA 22203-1954

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release,
distribution unlimited

20040130 056

13. SUPPLEMENTARY NOTES
DODAAD CODE: 80127
AFOSR Program Manager: Dr. Robert Herklotz

14. ABSTRACT
The objective of this project was to develop stegaiialysis techniques for images that have been potentially subjected to a xvatermarking algorithm. Our effort was directed towards two separate research areas: Detection of robust watennarking techniques using quality measures and detection of fragile authentication watermarks.

In the first part of the proposed research our objective was to develop universal steganalysis techniques for identifying the presence of robust digital watermarks by using image quality metrics to detect artifacts induced into the image by the watennarking process. Specific objectives included:

I. Identification of image quality features that best aid in differentiating between watemmrked and non-watermarked images
2. The development of classification techniques that separate such images in the selected feature space.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr Nasir Menon
					19b. TELEPHONE NUMBER (Include area code) 718-260-3970

Final Report
Steganalysis of Digital Watermarking Techniques

AFOSR Contract F49620-01-1-0243
January 2001 to December 2002.

Nasir Memon
Polytechnic University
Brooklyn, NY 11201
memon@poly.edu
(718)-260-3970

Jessica Fridrich
SUNY Binghamton
Binghamton, NY 13902
fridrich@binghamton.edu
(607) 777-2577

1. Cover Sheet: Attached.

2. Objectives: The objective of this project was to develop steganalysis techniques for images that have been potentially subjected to a watermarking algorithm. Our effort was directed towards two separate research areas: Detection of robust watermarking techniques using quality measures and detection of fragile authentication watermarks.

In the first part of the proposed research our objective was to develop universal steganalysis techniques for identifying the presence of robust digital watermarks by using image quality metrics to detect artifacts induced into the image by the watermarking process. Specific objectives included:

1. Identification of image quality features that best aid in differentiating between watermarked and non-watermarked images
2. The development of classification techniques that separate such images in the selected feature space.

In the second part of the proposed research we focused on detection of images watermarked with authentication watermarks for image integrity protection. Because authentication watermarks are almost always very weak signals, the methodology cannot be based on quality metrics.
The project was done as a collaborative effort between the PI Nasir Memon at Polytechnic University and the Co-PI Jiri Fridrich at SUNY Binghamton. Project deliverables include a working prototype software for steganalysis, implemented on a PC platform that was delivered to AFRL for experimentation.

3. Status of Effort

Our objectives towards development of steganalysis techniques based on image quality metrics were more than met. We conducted extensive experimentation with a large image data set to identify image quality features that best aid in differentiating between watermarked and non-watermarked images. We then developed classification techniques that separate such images in the selected feature space. Results were published in two conference papers and one journal paper.

Our work has made good progress in taking us towards developing universal steganalysis techniques that can distinguish between cover objects and stego objects. It representing the first published effort that such a technique is indeed possible. However, we need further improvements in the false positive and miss rates to make the techniques more robust and reliable.

Besides steganalysis based on image quality metrics, we also developed a steganalysis technique based on binary similarity metrics. This gave results comparable to those obtained by image quality metrics but at a significantly reduced cost in terms of computational complexity.

Our work in steganalysis led us top the question as to how much data can be embedded in an image without it being reliably detected? We formulated the problem in a theoretic setting and have given some preliminary answers. These answers provide some insight on the fundamental bounds of steganalysis. We are currently studying the problem further to better understand these bounds.

4. Summary of Achievements

We made progress on several fronts in the project. Below we itemize these achievements by topic and summarize the main results obtained. More detailed results are in the papers listed in the publications section.

1. ***Image Quality Metric (IQM) Based Steganalysis:*** Our earlier preliminary work had suggested that a particular watermarking scheme leaves statistical evidence or structure that can be exploited for detection with the aid of proper selection of image features and multivariate regression analysis. In this work, we identified some sophisticated image quality metrics as the feature set to distinguish between watermarked and unwatermarked images. To identify specific quality measures, which provide the best discriminative power, we used analysis of variance (ANOVA) techniques. We conducted extensive experiments with different feature sets and classification techniques using well-known and commercially available watermarking techniques. The results obtained validate our approach and we were able to distinguish between watermarked and unwatermarked images with moderate accuracy. However, a significant amount of further experimental work and mathematical analysis is needed before we get a better understanding about the nature

of artifacts caused by watermarking and the best means to exploit this knowledge for the purpose of steganalysis. Initial results of our technique were presented in [19]. Additional results showing the ability to detect unknown steganographic algorithm was presented in [17]. Finally, these results were consolidated in one journal paper in [1].

2. ***Binary Similarity Measure Based Steganalysis:*** One of the limitations of IQM based steganalysis was the computation time needed. Hence we started looking at computationally simpler approached. Our efforts led us to the development of a novel technique for that employs the seventh and eight LSB's in an image to compute a set of binary similarity measures. The basic idea is that, there must be more correlation in these bits in a clean image than in a stego-image, as the 8th bit in a stego image is relatively random. The steganalyzer, that is, the marked – non-marked classifier, is built using multivariate regression on the set of computed binary similarity measures. Simulation results with a set of images and well-known LSB type steganographic techniques indicate that the new steganalyzer provides promising results. One way to interpret this steganalyzer is that it uses binary similarity measures as image quality metrics for the purpose of steganalysis. Since these are simpler to compute, it leads to a more efficient IQM based steganalysis technique. A conference paper describing our approach and the results was written [10]. A journal paper is currently under preparation.

3. ***LSB Steganalysis for Images:*** We also developed another LSB steganalysis technique that can detect the existence of hidden messages that are randomly embedded in the least significant bits of natural continuous-tone images. The technique is inspired by the RS-Steganalysis technique of et al. and just like RS-Steganalysis, it can also precisely measure the length of the embedded message, even when the hidden message is very short relative to the image size. The key to our success is the formation of some subsets of pixels whose cardinalities change with LSB embedding, and such changes can be precisely quantified under the assumption that the embedded bits are randomly scattered. Interestingly, our study on steganalysis of LSB embedding sheds light on the recent work of Fridrich et al. on the detection of LSB embedding, and offers an analytical proof of an observation made by them. A preliminary paper describing our approach was presented at ICIP [11]. Additional investigation using this approach is being currently carried out. This part of the research was not originally planned in the proposal but just developed due to our activity in steganography and discussion with other researchers, namely Xiaolin Wu and his post-doctoral student Sorina Dumetrescu.

4. ***Mathematical Analysis of Steganographic Capacity:*** During our research it occurred to us that although there have been many techniques for hiding messages in images, there has been little mathematical analysis establishing their statistical indistinguishibility from cover images. Hence we started looking at some specific image based steganography techniques and derived a closed form expression of the probability of false detection in terms of the number of bits that are hidden. This led us to the notion of steganographic capacity, that is, how many bits can we hide in a message without causing statistically significant modifications? Our results are able

to provide an upper bound on this capacity. This work was done in collaboration with R. Chandramouli of Stevens Institute and published in ICIP 2001 [8]. We believe this is a promising area of work and will improve our mathematical understanding of steganography and steganalysis. Again, this was work not originally planned in the proposal but developed as a by product of our activity in steganography.

5. ***Audio Steganalysis:*** We had always maintained our quality metric based steganalysis techniques for images were equally applicable to audio and video, with appropriate selection of corresponding quality measures. A PhD student of our NSF collaborator on this project, B. Sankur has started looking at statistical methods to detect the presence of hidden messages in audio signals. Initial experimental results show that the proposed technique can be used to detect the presence of hidden messages in digital audio data. These results were published in [7].

6. ***Covert Channels by Data Masking:*** While pursuing our investigation into steganography and steganalysis, it occurred to us that the basic model of steganography can be turned on its head to produce interesting alternative techniques. Steganography strives to embed a secret message into a cover object in such way that the cover object and stego object are statistically and perceptually indistinguishable. However, in any automated system, perceptual tests are not practical and only statistical tests will be carried out. This means that given a secret message, we could "massage" it into a stego object that looks like to cover object from a statistical point of view and not from a perceptual point if view. Using audio stego objects we could show that an order of magnitude additional bits can Results were published in [9].

7. ***Other Work:*** In addition to steganalysis also worked on related problems. These include:

    - ***Image Authentication:*** We analyzed a well known robust hash function for image data called the Visual Hash Function (VHF) and showed that it is susceptible to attacks. Given just an input and its hash value, we showed how to construct a statistical model of the hash function, without any knowledge of the secret key used to compute the hash. This model can then be used to engineer arbitrary and malicious collisions. We proposed a possible modification to VHF so that constructing a model that mimics its behavior becomes difficult. Results were published in [6].

    - ***Image Reassembly***: Reassembly of fragmented objects from a collection of randomly mixed fragments is a common problem in classical forensics. We addressed the digital forensic equivalent, i.e., reassembly of document fragments, using statistical modeling tools applied in data compression. We showed how we can recover images and documents, only from their pieces. Results were published in [12].

- ***Watermarking Protocols:*** In [4] we presented a watermarking that enables Alice to demonstrate the presence of watermark to Bob without revealing the watermark. This is a difficult problem and although our proposed protocol was not a zero-knowledge protocol, it was the first step in the future development of such a protocol.

5. Personnel Supported:

Nasir Memon, I. Avcibas, B. Sankur, Kulesh Shanmugasundaram, and Mike Sosonkin.

6. Publications

1. I. Avcibas, N. Memon, B. Sankur. Steganalysis using Image Quality Metrics. *IEEE Transactions on Image Processing*, 12(1): 221-229, February 2003.
2. R. Chandramouli and N. Memon. On Sequential Watermark Detection. To appear in *IEEE Transactions on Signal Processing.* 51 (4):1034-1044, April 2003.
3. R. Radhakrishnan, N. Memon. On the Security of the SARI Image Authentication System. IEEE *Transactions on Circuits and Systems for Video Technology*, 12(11):1030-1033, November 2002.
4. K. Gopalakrishnan, N. Memon and P. Vora. Protocols for Watermark Verification. IEEE Multimedia, 8(4):66-70, October 2001.
5. J. Fridrich, M. Goljan and N. Memon. On the Security of the Yeung-Mintzer Fragile Watermarking Scheme. *Journal of Electronic Imaging*, 11(02), 262-274, April 2002.
6. R. Radhakrishnan, Z. Xiong, N. Memon. On the Security of Visual Hash Function. *Security and Watermarking of Multimedia Contents*, San Jose, CA, January 2003.
7. I. Avcibas, N. D. Memon, B. Sankur, Y. Yigit, O. Kahya. Audio steganalysis with statistical distance metrics. *Security and Watermarking of Multimedia Contents*, San Jose, CA, January 2003.
8. R. Chandramouli, N. Memon. Steganography capacity: a steganalysis perspective. *Security and Watermarking of Multimedia Contents*, San Jose, CA, January 2003.
9. R. Radhakrishnan, K. Shanmugasundaram and N. Memon. Data Masking: A Secure-Covert Channel Paradigm. *IEEE Workshop on Multimedia*, St. Thomas, Virgin Islands, December 2002.
10. I. Avcibas, N. Memon, B. Sankur. Image Steganalysis with Binary Similarity Measures. *IEEE International Conference on Image Processing*, Rochester, New York, September 2002.
11. S. Dumitrescu, X. Wu, N. Memon. On Steganalysis of Random LSB Embedding in Continuous-tone Images. *IEEE International Conference on Image Processing*, Rochester, New York, September 2002.
12. K. Shanmugasundaram and N. Memon. Automatic Reassembly of Document Fragments via Data Compression. *Digital Forensics Research Workshop*, Syracuse, NY, August 2002.

13. R. Chandramouli, G. Li and N. Memon. Adaptive Steganography. *Security and Watermarking of Multimedia Contents*, San Jose, CA, February 2002.
14. R. Radhakrishnan and N. Memon. Audio Content Authentication Based on Psycho-Acoustic Model. *Security and Watermarking of Multimedia Contents*, San Jose, CA, February 2002.
15. R. Chandramouli and N. Memon. Analysis of LSB-based Image Steganography techniques. *IEEE International Conference on Image Processing*, Thessaloniki, Greece, October 2001.
16. R. Regunathan and N. Memon. On the Security of the SARI Image Authentication System. *IEEE International Conference on Image Processing*, Thessaloniki, Greece, October 2001.
17. I. Avcibas, N. Memon and B. Sankur. Steganalysis based on Image Quality Metrics - Differentiating between techniques. *IEEE Workshop on Multimedia*, Cannes, France, October 2001.
18. M. Chen, E. Wong, N. Memon and S. Adams. Recent Developments in Document Image Watermarking and Data Hiding. *Multimedia Systems and Applications, SPIE OPTICOM*, Denver, August 2001.
19. I. Avcibas, N. Memon and B. Sankur. Steganalysis using Image Quality Metrics. *Security and Watermarking of Multimedia Contents*, San Jose, CA, February 2001.

7. Interactions/Transitions:

1. Presented papers at ICIP 2001 and 2002. SPIE 2001 and 2002. MMSP Workshop 2001 and 2002.
2. Presentation to site visit by AFRL in November 2002.
3. Panel member, MMSP panel on "Future of Watermarking". MMSP 2002.

8. New discoveries, inventions, or patent disclosures:

We will explore the possibility of patenting our image reassembly work and steganalysis work with AFRL. Possibilities were discussed during our meetings.|

9. Honors/Awards: None.

10. Markings: None.